Assessment Submission Cover Sheet

This Assessment Cover Sheet **must** be included on all Assessment submissions.

|  |  |
| --- | --- |
| Assignment Title | Assignment B – Portfolio Assessment |
| Module | Data Mining |
| Student Name  (same as Student Card) | Ciaran Finnegan |
| Student Number |  |
| Programme |  |
| Part-Time/Full-Time |  |
| Year of Study  (First Year, Second Year, etc) |  |

Late Submissions: Assessment submitted after the deadline will have a late penalty applied.

**Academic Integrity for assessment in TU Dublin Programmes**

Each student is responsible for knowing and abiding by TU Dublin Academic Regulations and Policies. Any student in breach of these regulation/policies will be subject to action in accordance with the University’s procedures for breaches of assessment regulations. Please refer to the General Assessment Regulations at

<https://tudublin.libguides.com/c.php?g=674049&p=4794713>

<https://www.tudublinsu.ie/advice/exams/breachesofregulations/>

All students are expected to complete their courses/programmes in compliance with University regulations. No student shall engage in any activity that involves attempting to receive a grade by means other than honest effort, for example:

1. No student shall complete, in part or in total, any examination or assessment for another person.
2. No student shall knowingly allow any examination or assessment to be completed, in part or in total, for themselves by another person.
3. No student shall plagiarise or copy the work of another and submit it as their own work.
4. No student shall falsify any data. Falsification is the invention of data, its alteration, its copying from any other source, or otherwise obtaining it by unfair means, or inventing quotations and/or references.
5. No student shall use aids or devices excluded by the lecturer in undertaking course work or assessments/ examinations.
6. No student shall knowingly procure, provide, or accept any materials that contain questions or answers to any examination or assessment to be given at a subsequent time.
7. No student shall provide their assignments, in part or in total, to any other student in current or future classes of this module/ programme unless authorised to do so by the lecturer.
8. No student shall submit substantially the same material in more than one module/programme without prior authorization.
9. No student shall alter graded assignments or examinations and then resubmit them for regrading, unless specifically authorised to do so by the lecturer.
10. All programming code and documentation, unless correctly referenced, submitted for assessment or existing in the student’s computer accounts must be the students’ original work or material specifically authorized by the lecturer.
11. Collaborating with other students to develop, complete or correct course work is limited to activities explicitly authorized by the lecturer.
12. For all group assignments, each member of the group is responsible for the academic integrity of the entire submission. Consequently, all group members must satisfy themselves that all elements of their submission adhere to the academic integrity statement points above.

By submitting coursework, either physically or electronically, you are confirming that it is your own work (or, in the case of a group submission, that it is the result of joint work undertaken by members of the group that you represent) and that you have read and understand the University’s Regulations and Policies covering Academic Integrity (see General Assessment Regulations)*.*

Coursework may be submitted to an electronic detection system in order to help ascertain if any plagiarised material is present. If you have queries about what constitutes plagiarism, please speak to your lecturer.

|  |  |
| --- | --- |
| Student Signature |  |
| Date |  |

IMPORTANT:

* Complete the required number of tasks as defined in Assessment Handout
* The sections listed below are an example of the section headings for each task. You can use alternative headings
* Tasks 1-3: Sub-Sections 1-7 should be no longer than 8 pages (minimum 6 pages), including diagrams, images, screen captures, tables, etc. Careful selection of these is needed.
  + Code does not count to this total. Code should be added to the relevant section.
* Detailed discussion is expected. Marks are awarded based on depth of information given.
* Marks are awarded based on complexity of problem and depth of work.

# TASK 4 – *Ethics and the user of Data Science/ML/AI*

## Task 4-1 : Stop The Killer Robots – Autonomous Drone Warfare

1. **Overview of problem**

[Robots Aren't Better Soldiers than Humans (hrw.org)](https://www.hrw.org/node/376854/printable/print)

For years, roboticist [Noel Sharkey, a professor at Sheffield University in England, warned that computers may be better than humans at some tasks](https://www.unog.ch/80256EDD006B8954/(httpAssets)/78C4807FEE4C27E5C1257CD700611800/$file/Sharkey_MX_LAWS_technical_2014.pdf), but killing is not one of them. Sharkey and his colleagues became increasingly alarmed that technological advances in computer programming and sensors would make it possible to develop systems capable of selecting targets and firing on them without human control.[ref]

We already have examples of quasi-autonomous weapons in the field, such as Israel’s anti-radar "fire and forget" drone ..all of which have raised a debate around the moral dilemma of such technology [x]

Autonomous weapons systems would be more likely to carry out unlawful orders if programmed to do so, due to their lack of emotion and the fact that morality cannot be outsourced to machines.

1. **Ethical and Legal Challenges**

The MoD says its policy is that only humans will be able to fire weapons, although there is growing concern about the potential danger of unrestricted robot warfare, led by the Campaign to Stop Killer Robots.

Technology under development includes the i9 drone, which is powered by six rotors and carries two shotguns. Remotely operated, it is intended to be used to storm buildings, typically an urban warfare situation that generates some of the highest casualties.

the Pentagon faces the daunting challenge of working through a plethora of ethical issues surrounding the technology while staying ahead of advanced adversaries who are pursuing their own capabilities.

An effective ‘AI Arms Race’. See reflections…

In August, heads were turned when an AI agent defeated a seasoned F-16 fighter pilot in a series of simulated combat engagements during the final round of the Defense Advanced Research Projects Agency’s “Alpha Dogfight” Trials. The agent, developed by Heron Systems, went undefeated with a record of 5-0 against the airman whose call sign was “Banger.”

“It’s a significant moment,” said Peter W. Singer, a strategist and senior fellow at the New America think tank, comparing it to chess master Garry Kasparov losing to IBM’s Deep Blue computer at the complex game.

During the simulated dogfight “the AI shifted [its tactics] and it kept grinding away in different ways at him” until it won, noted Singer, co-author of *Ghost Fleet* and *Burn-In*, which examine the military and societal implications of autonomy and artificial intelligence.

Retired US Colonel Brian Hall, an Autonomy Program Analyst at the Joint Chiefs of Staff wrote in 2017 [ref] that the advantage of autonomous weapons systems will "come from augmenting human decision making, not replacing it."

1. **Challenges for Data Scientist**

The principles need to be kept in mind throughout the acquisition process and product lifecycle when officials are thinking about how to design, develop, deploy and use AI, she said.

“We’re currently working with RDT&E folks in terms of thinking through how we can integrate the ethics aspects in their test harness” for software and other technology, Patel said. “We’re looking at the testing aspects, the algorithmic aspects, the system integration, and then the human-machine teaming aspects. … All of those pieces are critical aspects or potential areas for us to embed and engage in from a responsible AI perspective.”

Developers are making strides in AI, adding urgency to the department’s efforts to craft new policies for the ethical deployment of the capabilities. In August, heads were turned when an AI agent defeated a seasoned F-16 fighter pilot in a series of simulated combat engagements during the final round of the Defense Advanced Research Projects Agency’s “Alpha Dogfight” Trials. The agent, developed by Heron Systems, went undefeated with a record of 5-0 against the airman whose call sign was “Banger.”

“It’s a significant moment,” said Peter W. Singer, a strategist and senior fellow at the New America think tank, comparing it to chess master Garry Kasparov losing to IBM’s Deep Blue computer at the complex game.

Artificial intelligence must work as intended, or else bad things could happen and users won’t trust it. One issue that could undermine trust is known as algorithmic bias.

“Algorithmic bias is basically when either [the system] was trained in the wrong way for a scenario that it was applied to, or it was provided biased data of some kind,” Singer explained.

For example, in the civilian world there was a case where an artificial intelligence tool was used to aid in the treatment of heart disease, but it was providing bad medical advice for African Americans. “No one told that AI, ‘You be racist,’” Singer said. “But it was, because of the way it was trained in the data.”

Paul Scharre, director of the Technology and National Security Program at the Center for a New American Security and the author of *Army of None: Autonomous Weapons and the Future of War*, said the laws of armed conflict have long been baked into how the Pentagon incorporates new technology. But artificial intelligence isn’t like standard weapon systems, and it requires more oversight.

“What I think you’ve seen DoD do, which I think is the right step, is say, ‘AI seems to have something different about it,’” Scharre said. “Because of how it changes the relationship with humans and human responsibility for activity, because of some of the features of the technology today and concerns about … reliability and robustness, we need to pay more attention to AI than we might normally would to, say, a more advanced missile or some other kinds of technology.”

“That means that we have two kinds of legal and ethical questions that we’ve really never wrestled with before. The first is machine permissibility. What is the tool allowed to do on its own? The second is machine accountability. Who takes responsibility … for what the tool does on its own?”

Noel Sharkey says..

“Their usewould raise a substantial accountability gap when it comes to removing human control from the use of force, finding that programmers, manufacturers, and military personnel could all escape liability for unlawful deaths and injuries caused by [fully autonomous weapons](https://urldefense.com/v3/__https:/www.hrw.org/topic/arms/killer-robots__;!!BspMT6SJLSDJ!bzJHVgHJIi0SAQy_DQ7gqLvex1vGQgFeqL6uEdwkSzJaxpX-6pxINnn0vF9HLUHwcuYR_Q$).”

In February 2020, the Defense Department rolled out a list of five AI ethical principles based on recommendations from the Defense Innovation Board and other experts inside and outside of the government.[ref]

Military personnel must be responsible and exercise appropriate levels of judgment and care while remaining responsible for the development, deployment and use of AI capabilities, according to the list.

The technology should be “equitable” and steps taken to minimize unintended bias.

It must be traceable: “The department’s AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation,” according to the list.

Systems must also be reliable: “The department’s AI capabilities will have explicit, well-defined uses, and the safety, security and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire lifecycles.”

And finally, they must be governable: “The department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.”

[Artificial intelligence arms race - Wikipedia](https://en.wikipedia.org/wiki/Artificial_intelligence_arms_race)

Lots of references in above link – maybe enough to use alone

Will lead to corners being cut…

Scharre noted that AI is also vulnerable to hacking or spoofing attacks that could corrupt data or cause other problems.

The Pentagon will have to continually make decisions about which tasks to delegate to machines, he noted. A key question will be where humans will be in the decision-making loop.

The Pentagon’s Joint Artificial Intelligence Center has been tasked with developing the policies to turn the new AI principles into practice, and leading implementation across the department. Its policy recommendations will be delivered to leadership by the end of this year, according to Alka Patel, head of AI ethics policy at JAIC.

Alka Patel, the center’s head of AI ethics policy, told FedScoop her goal is to make “ethics a natural part of the way we all think.”[ref]

One of the ways JAIC officials think the DOD can lead in the ethical application of AI is through strong testing and evaluation, said Patel and Jane Pinelis, the center’s head of testing and evaluation. JAIC officials have predicted that the Pentagon will be a global leader in the testing and evaluation of AI due to the seriousness of its initiatives.

“We have tried to build ethics into every piece of the test and evaluation process,” said Pinelis, who participated in the pilot cohort for the Responsible AI Champion’s program. Few programs in the JAIC have reached the level of maturity to run through all the different types of testing, she added.

JAIC acting Director Nand Mulchandani has stressed he wants to see more testing tools come from the private sector.

“A lot of AI testing is being done manually,” he said last week. There are “not enough tools and products for testing.”

Patel noted that implementation policies will need to ensure that the technology has disengagement mechanisms in place in case something goes wrong.

“In many cases when we think about implementation of those principles themselves, they really speak to good engineering practices in terms of capability, in terms of reliability, in terms of governability,” she said.

The center is also engaging with industry and academia — organizations that will help design and build the systems — as it develops policies, to include requests for information.

JAIC is also engaging other nations as it works through policy issues. The center in September hosted the first-ever AI Partnership for Defense with military officials from 13 countries. The two-day meeting focused on ethical principles and best practices for implementation.

Experts say the world won’t be overrun by godless killing machines anytime soon. There are more pressing concerns, such as working through shortcomings in artificial intelligence and machine learning, and figuring out how best to conduct human-machine teaming.

Mr. Rickli said drones today were being used in swarms, particularly in the civilian sector. Humans could not control swarming drones, this had to be done by a computer. In the military, there was the principle of mass fire power speed and concentration, and there was already some manifestation of this. In 2017, ISIS used a commercial drone that they weaponised with small hand grenades and for the first time, a non-State actor group managed to reach tactical air supremacy against the Iraqi armed forces. There had been other examples of attacks using swarm drones since then. Research was being invested by major powers into swarming. The fastest development would be in the cyber domain, because there were less obstacles. As for the impact of all this on weapons of mass destruction, if drones and algorithms of swarming were combined, the scalability was enormous, and more so in the cyber domain. It was impossible to find a system that could defend against the major threat that swarms of drones represented. This could favour the international arms race. Research had shown that the best results were made when humans and machines cooperated together. In the military, research was being conducted for a pilot to be able, thanks to artificial intelligence, to coordinate a swarm of drones. These concepts were called loyal wingmen, and most powers were investing in this technology.

[CONFERENCE ON DISARMAMENT HOLDS THEMATIC PLENARY DISCUSSION ON NEW WEAPONS OF MASS DESTRUCTION | UN GENEVA](https://www.ungeneva.org/en/news-media/meeting-summary/2020/09/la-conference-du-desarmement-discute-des-nouvelles-armes-de)

Concern… None of the nine United Nations meetings held from 2014 to 2020on killer robots have focused at any length on how better programming could be the solution. There remains a lack of interest in discussing whether there are potential benefits or advantages to removing meaningful human control from the use of force. This shows how technical fixes proposed years ago are, on their own, not an adequate or appropriate regulatory response.

1. **Reflections**

Singer said the Pentagon might someday loosen the restrictions it has placed on using artificial intelligence. An historical analogy is the U.S. Navy’s embrace of unrestricted submarine warfare in World War II. Moral opposition to the German use of the tactic against civilian vessels was a catalyst for the U.S. entry into World War I, he noted. But just a few hours after the Pearl Harbor attack in 1941, the order went out to wage unrestricted submarine warfare against Japan.

Could providing more autonomy in weapons systems result in greater accuracy and precision? Could such weapons increase compliance with international humanitarian laws because they would not rape or commit other war crimes? Would they “perform more ethically than human soldiers,” as one roboticist [claimed](https://www.unog.ch/80256EDD006B8954/(httpAssets)/FD01CB0025020DDFC1257CD70060EA38/$file/Arkin_LAWS_technical_2014.pdf)?

[Artificial intelligence arms race - Wikipedia](https://en.wikipedia.org/wiki/Artificial_intelligence_arms_race)

Lots of references in above link – maybe enough to use alone

Will lead to corners being cut…

With military investments in artificial intelligence and emerging technologies increasing unabated, Sharkey and his colleagues demanded arms control. Yet China, Israel, Russia, South Korea, Britain, the United States, and other military powers have continued their development of air, land, and sea-based autonomous weapons systems.

New global treaties – comparable to nuclear treaties of the past…

A new international treaty to prohibit and restrict killer robots has been endorsed by [dozens of countries](https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and) [ref]..but the major powers are resistant to this idea, preferring to look at existing legislation regulation [ref][ref]

.

The Conference on Disarmament and the General Assembly had discussed many times the necessity of a universal, legally binding treaty to prohibit the development, production, stockpile and use of new types of weapons of mass destruction

As Russian President Vladimir Putin said in September, whoever is the leader in artificial intelligence "will become the ruler of the world."

While the powerful potential of autonomous weapons on the battlefield causes concerns, it also makes them more difficult to ban or regulate, experts said.

"This is the arms control dilemma. The more useful potential weapons are for militaries, the harder it is to regulate or ban them," Horowitz said. "Uncertainty about what an autonomous weapon is further complicates the discussion – states are unlikely to agree to regulations or bans if they do not know what will be covered," he added.

For Franke, an outright ban or arms-control regime is unlikely. Lethal autonomous weapons systems are not like nuclear weapons since they cannot be counted, which is a key requirement for arms control agreements. They also are unlike chemical weapons, which have been banned. And with no strict definition of what a lethal autonomous weapons system is, "there is no way to identify it by just looking at it," she said.

However, should we not try…?

Is an outright ban possible? Many in the military today believe that is impossible [ref-no turning back]

The Stop TKR group mainly advocates an outright ban. However, there are voices in the organisation, with both academic and military backgrounds, that push for governments to, at the very least, adopt the use of ethical principles in AI weapons development and enshrine this process in multi-lateral treat agreements. Critically, it should be paramount that ‘permanent significant human control’ remains in place at all time [Argy ref]

1. **References**

Use one of the commonly used References and Citation formats.

[The-evolution-of-disruptive-technologies-and.pdf (stopkillerrobots.org)](https://www.stopkillerrobots.org/wp-content/uploads/2021/09/The-evolution-of-disruptive-technologies-and.pdf)

[Stop Killer Robots](https://www.stopkillerrobots.org/)

## Task 4-1 : <Title of Case Study)

1. **Overview of problem**
2. **Ethical and Legal Challenges**
3. **Challenges for Data Scientist**
4. **Reflections**
5. **References**

Use one of the commonly used References and Citation formats.